PharmaSUG 2018 - SDE NORTH CAROLINA

ADaM Structures for Integration: A Preview

Wayne Zhong, Accretion Softworks; Kimberly Minkalis, The Griesser Group;
Deborah Bauer, Sanofi;

ABSTRACT

Integration and analysis of data across all studies in a submission is a vital part of applications for regulatory approval in the pharma industry. The existing ADaM classes (ADSL, BDS, and OCCDS) already support some simple cases of integration analysis. However, there has been a need for an integration standard that supports the more complex cases.

To address this need, the ADaM Integration sub-team is developing the upcoming ADaM Integration standards document. This paper introduces the new IADSL, IBDS, and IOCCDS classes found in this document. IADSL allows for multiple records per subject. IBDS and IOCCDS work effectively with the new IADSL class. This paper also discusses the analysis needs that necessitated the creation of the new classes, and provides examples in the form of usage scenarios, data, and metadata. With them, no future integration will prove too complex.

INTRODUCTION

Integrated clinical data is data collected from multiple clinical studies, combined together in a consistent manner, to support safety and efficacy analyses as required by regulatory agencies. Other uses include annual safety reports, investigator brochures (IB), ongoing safety monitoring and risk management, marketing materials, integrated PK analyses, extension studies, and answers to regulatory authorities' questions.

Particularly, the FDA requires an Integrated Summary of Safety (ISS) and an Integrated Summary of Efficacy (ISE) and expects that "the ISE and ISS are not summaries but rather detailed integrated analyses of all relevant data from the clinical study reports that belong in Module 5." [1] Because ISS and ISE analysis data are critical components of pharmaceutical and biotech applications to the FDA, the ADaM integrated data structures were developed with this purpose in mind.

This paper describes two approaches for preparing integrated ADaM data: 1) In cases where the producer determines that using existing ADaM classes (ADSL, BDS, OCCDS) will support analysis, provide clear metadata, and offer traceability back to source data, then the existing classes may be used. 2) For all other cases, the ADaM Integration sub-team has prepared new data structure classes (IADSL, IBDS, IOCCDS) to support the unique challenges of integration analysis. This paper will introduce and give examples of the use of these new classes. Please note that the ADaM integration sub-team does not require or recommend a given data flow when creating integrated ADaM dataset.

This paper assumes the reader has working knowledge of CDSIC ADaM standards and experience with integration analysis, specifically integration pool definitions and analyses based on pools.

INTEGRATION WITH EXISTING STRUCTURES

In cases where the producer determines that using existing ADaM classes (ADSL, BDS, OCCDS in Figure 1) will support analysis, provide clear metadata, and offer traceability back to source data, then the existing classes may be used.

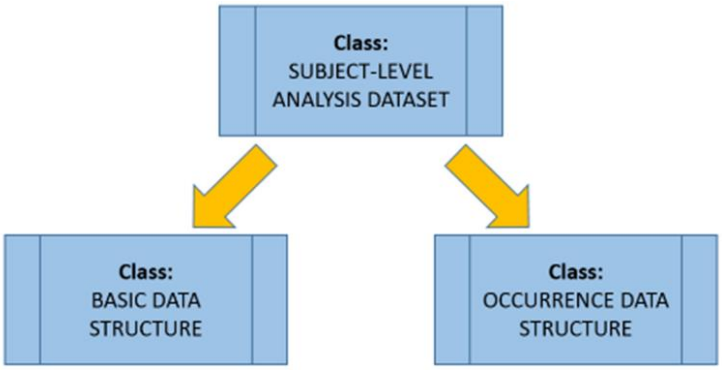

Figure 1: Existing ADaM Structure Classes

In the simple case, where each subject in the integration participates in only one study, the study-level ADSL datasets may be set together to create an integrated one-record-per-subject ADSL dataset with a class of ADSL. BDS and OCCDS study-level data may similarly be combined.

In complex cases, where subjects participate in more than one trial, existing classes may still be used. However, existing rules must still be followed, e.g. required ADSL variables such as STUDYID, SUBJID must still be populated, which means choosing one value for subjects participating in more than one study.

If the integrated SAP defines pool analysis with treatment, baseline, covariate values that vary for a subject depending on the pool, variables to support these differences will be needed. E.g. if the first exposure date TRTSDT differs for a subject depending on the pool, then pool-specific versions of this variable will be needed. Due to the breadth of variables that can be impacted by requirements unique to integration, variable naming is left to the sponsor; standardized variable naming for integration analysis will be a part of the approach using new integration structure classes.

EXAMPLE 1: INTEGRATED SUMMARY OF EFFICACY

In this example, efficacy data from three double-blind Phase 3 studies (S001, S002, S003) will be combined to create Integrated Summary of Efficacy (ISE) data.

Study S001 contained a placebo run-in period and a double-blind period, while studies S002 and S003 contained one double-blind period. The study-level ADaM datasets were created with standardized analysis variables and controlled terminology across the three studies with future integration in mind. As a result, the integrated ADSL and integrated ADEFF datasets can be created by stacking the study ADaM datasets (Figure 2). Some additional efforts, such as re-deriving treatment variables to align the double-blind period across the three studies (Table 1) and re-deriving the numeric order variable will be necessary. The integrated ADSL datasets has a class of "SUBJECT LEVEL ANALYSIS DATASET" and ADEFF has a class of “BASIC DATA STRUCTURE”.

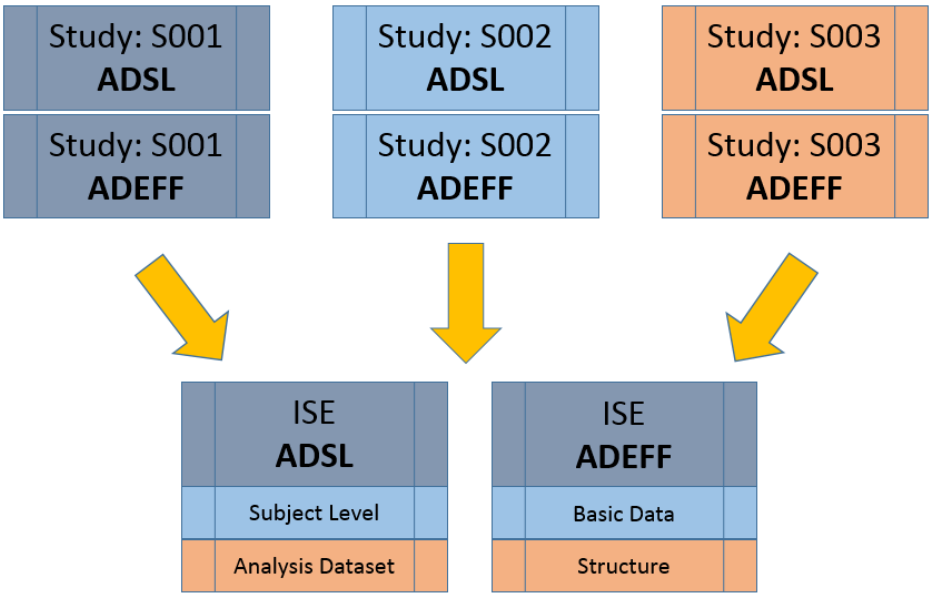

Figure 2: Data Flow for ISE Example

The primary efficacy analysis in this ISE is a change from baseline to endpoint for total PANSS score for all randomized treatment groups in the three studies. Each study medication (Drug A) dose group will be compared with Placebo using an analysis of covariance model (ANCOVA) based on the pooled ITT analysis population excluding the active control group (Drug B 10 mg). Sample metadata and data are presented below.

ADSL

The ADSL metadata in Table 1 below describes the derivation and data origin for variables in the integrated ADSL dataset. Please note that the double-blind period for study S001 has been aligned with the other two studies. The actual integrated ADSL dataset will contain many more variables; this example has been edited for length for this paper.

Table 1: Sample ADSL Metadata

Variable	Where Condition	Source/Derivation/Comments
STUDYID		ADSL.STUDYID from individual study ADSL dataset
USUBJID		ADSL.USUBJID from individual study ADSL dataset
SITEID		ADSL.SITEID from individual study ADSL dataset
ARMCD		ADSL.ARMCD from individual study ADSL dataset
TRT01P	STUDYID= “S001”	S001.ADSL.TRT02P
-	STUDYID in (“S002”, “S003”)	S002.ADSL.TRT01P and S003.ADSL.TRT01P
SAFFL		ADSL.SAFFL from individual study ADSL dataset
ITTFL		ADSL.ITTFL from individual study ADSL dataset
AGE		ADSL.AGE from individual study ADSL dataset
REGION1		ADSL.REGION1 from individual study ADSL dataset

Table 2 below shows sample ADSL records for the variables described in Table 1. Other than changes in the value for variable STUDYID, there is little to distinguish this dataset from a study-level ADSL dataset.

Table 2: Sample ADSL Data

ROW	STUDYID	USUBJID	SITEID	ARMCD	TRT01P	SAFFL	ITTFL	AGE	REGION1
1	S001	S001-001	0501	A16	Drug A 16 mg	Y	Y	44	North America
2	S001	S001-002	0501	PBO	Placebo	Y	Y	37	North America
3	S001	S001-003	0501	A8	Drug A 8 mg	Y	Y	66	North America
4	S001	S001-004	0503	A12	Drug A 12 mg	Y	Y	33	North America
5	S002	S002-001	0701	B10	Drug B 10 mg	Y	Y	65	Eastern Europe
6	S002	S002-002	0701	A8	Drug A 8 mg	Y	Y	43	Eastern Europe
7	S002	S002-003	0702	B16	Drug B 16 mg	Y	Y	38	Western Europe
8	S003	S003-001	0908	A8	Drug A 8 mg	Y	Y	55	North America
9	S003	S003-002	0908	A4	Drug A 4 mg	Y	Y	45	North America
10	S003	S003-003	0909	A12	Drug A 12 mg	Y	Y	67	North America
11	S003	S003-004	0909	B10	Drug B 10 mg	Y	Y	28	North America

ADEFF

The ADEFF metadata in Table 3 below describes the derivation and data origin for variables in the integrated ADEFF datasets. Please note that the TRTPN variable is rederived to ensure variables TRTP and TRTPN shared a one-to-one relationship as required by the BDS structure. The variables STUDYID and RSSEQ allow records to be traced back to the original SDTM source record.

Table 3: Sample ADEFF Metadata

Variable	Label	Source/Derivation/Comment
STUDYID	Study Identifier	ADEFF.STUDYID from individual study ADaM dataset
USUBJID	Unique Subject Identifier	ADEFF.USUBJID from individual study ADaM dataset
RSSEQ	Sequence Number	ADEFF.RSSEQ from individual study ADaM dataset
TRTP	Planned Treatment	ADEFF.TRTP from individual study ADaM dataset
TRTPN	Planned Treatment (N)	Mapped into unique number according to combined data for TRTP values.
PARAMCD	Parameter Code	ADEFF.PARAMCD from individual study ADaM dataset
ADY	Analysis Day	ADEFF.ADY from individual study ADaM dataset
VISITNUM	Visit Number	ADEFF.VISITNUM from individual study ADaM dataset
VISIT	Visit Name	ADEFF.VISIT from individual study ADaM dataset
AVISITN	Analysis Visit (N)	ADEFF.AVISITN from individual study ADaM dataset
AVISIT	Analysis Visit	ADEFF.AVISIT from individual study ADaM dataset

Variable	Label	Source/Derivation/Comment
AVAL	Analysis Value	ADEFF.AVAL from individual study ADaM dataset Derivation below is copied from individual study for convenience: - omitted for length -
BASE	Baseline Value	ADEFF.BASE from individual study ADaM dataset
CHG	Change from Baseline	ADEFF.CHG from individual study ADaM dataset
PCHG	Percent Change from Baseline	ADEFF.PCHG from individual study ADaM dataset
ABLFL	Baseline Record Flag	ADEFF.ABLFL from individual study ADaM dataset
ANL01FL	Analysis Record Flag 01	ADEFF.ANL01FL from individual study ADaM dataset (If multiple visits fall into the same visit window, then the one closest to the target day is chosen for analysis. These are flagged with ANL01FL="Y").
DTYPE	Derivation Type	ADEFF.DTYPE from individual study ADaM dataset. Value: LOV denotes that the last available observed value was used to impute the value for the given parameter and analysis visit.

Table 4 below shows sample ADEFF records for the variables described in Table 3. SDTM variables identify the records copied from SDTM, and derived records are copied directly from study ADaM resulting in a simple to produce ADaM datasets.

Table 4: Sample ADEFF Data

ROW	STUDYID	USUBJID	RSSEQ	TRTP	TRTPN	PARAMCD	ADY	VISITNUM
1	S001	S001-001	31	Drug A 16 mg	5	TOTPANSS	1	1
2	S001	S001-001	62	Drug A 16 mg	5	TOTPANSS	22	2
3	S001	S001-001	93	Drug A 16 mg	5	TOTPANSS	44	3
4	S001	S001-001		Drug A 16 mg	5	TOTPANSS	44	
5	S002	S002-001	31	Drug B 10 mg	6	TOTPANSS	-1	1
6	S002	S002-001	62	Drug B 10 mg	6	TOTPANSS	22	2
7	S002	S002-001		Drug B 10 mg	6	TOTPANSS	22	
8	S003	S003-001	31	Drug A 8 mg	3	TOTPANSS	-1	1
9	S003	S003-001	62	Drug A 8 mg	3	TOTPANSS	22	2
10	S003	S003-001	93	Drug A 8 mg	3	TOTPANSS	43	3
11	S003	S003-001		Drug A 8 mg	3	TOTPANSS	43	

ROW	VISIT	AVISITN	AVISIT	AVAL	BASE	CHG	PCHG	ABLFL	ANL01FL	DTYPE
1 (C)	BASELINE	1	BASELINE	83				Y	Y	
2 (C)	DAY 22	2	DAY 22	74	83	-9	-10.84		Y	
3 (C)	DAY 43	3	DAY 43	77	83	-6	-7.23		Y	
4 (C)		99	END POINT	77	83	-6	-7.23		Y	LOV
5 (C)	BASELINE	1	BASELINE	114				Y	Y	
6 (C)	DAY 22	2	DAY 22	95	114	-19	-16.67		Y	
7 (C)		99	END POINT	95	114	-19	-16.67		Y	LOV
8 (C)	BASELINE	1	BASELINE	88				Y	Y	
9 (C)	DAY 22	2	DAY 22	83	88	-5	-5.68		Y	
10 (C)	DAY 43	3	DAY 43	71	88	-17	-19.32		Y	
11 (C)		99	END POINT	71	88	-17	-19.32		Y	LOV

The simplicity of this example is possible due to both the straightforward design of the three pivotal studies and the effort to standardize the analysis datasets in the individual study analysis. Without either, the creation of the ISE analysis datasets would involve far more complexity.

INTEGRATION WITH NEW STRUCTURES

As an alternative for more complex integration analysis, the ADaM Integration sub-team has prepared new data structure classes (IADSL, IBDS, IOCCDS in Figure 3) to support the unique challenges of integration analysis. These three classes will be used together, i.e. if the IADSL class is used, then subsequent integrated ADaM datasets will need to use the IBDS and IOCCDS classes.

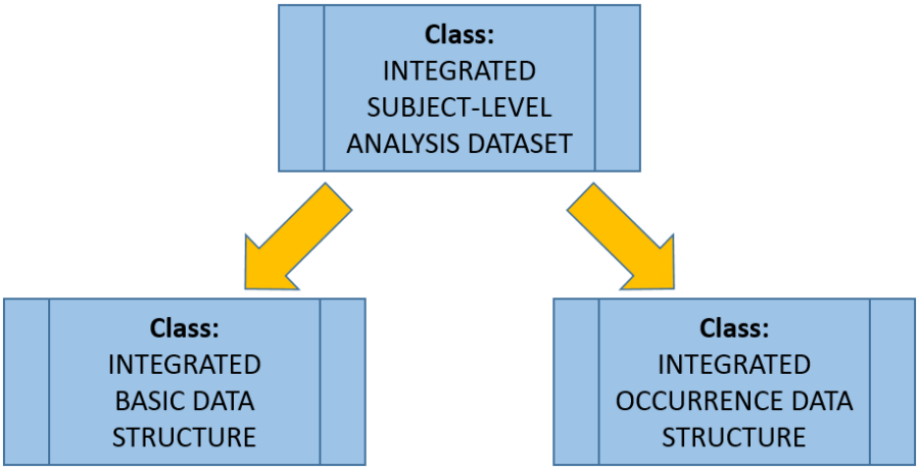

Figure 3: New ADaM Integration Structure Classes

The IADSL class permits multiple-records-per-subject, with the most basic structure being a one-record-per-subject-per-pool structure designed to support integrated SAPs that define treatment, baseline, or covariate values that vary for a subject depending on the pool. In these cases, subjects belonging to multiple pools will have one record created for each pool, and the value of the new standard variable POOL will follow the numbering convention defined in the SAP.

In taking this approach, the same variable in ADSL can take on a different value for each pool. E.g. if the first exposure date TRTSDT for a subject differs depending on the pool, the variable TRTSDT will hold the appropriate date in each pool-specific record.

The new IBDS and IOCCDS structure classes are designed for merging with the new IADSL structure. Where IADSL is a set of subject-level records for each pool, IBDS and IOCCDS is a set of records for each pool, allowing merging with ADSL by pool. Sets of records should only be created as needed for analysis, e.g. if a pool is not used for AE analysis, a set of records do not need to be created for that pool in ADAE. Records may also be omitted from a pool if it is not needed, e.g. if a pool only analyzes studies 2 and 3, records from study 1 do not need to be included for that pool.

The example below provides a demonstration of the new structure classes. Please note more integration ADaM standard variables and use cases may be found in the official ADaM Integration document.

EXAMPLE 2: INTEGRATED SUMMARY OF SAFETY

In this example, clinical data from five studies and multiple study phases will be combined to create the Integrated Summary of Safety (ISS). The study drug MD will be compared against the standard of care SOC. Integrated SDTM datasets will be used as the source data.

Study MD-101 is a dose escalation study on healthy subjects, MD-201 is a crossover study, MD-301/302 are pivotal double-blind studies, and MD-320 is an open-label extension study. Figure 4 shows the treatments received for three of subjects that progressed through these studies.

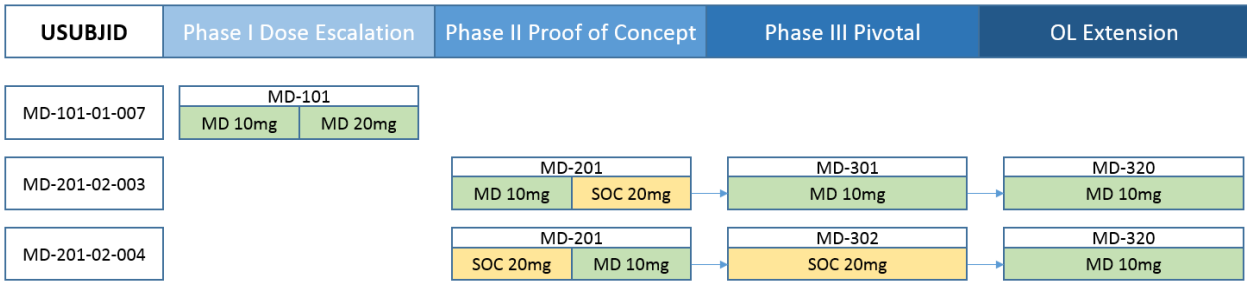

Figure 4: Study and Treatment Progression for Three Subjects

Figure 5 shows sample pool definitions from an integration SAP for the ISS analysis. Three pools are defined, listing the studies and treatment periods included for each pool.

Pool	Studies	Definition	Purpose
1	101, 201, 301, 302, 320	Overall Pool: Includes all periods.	Support treatment overview of all enrolled subjects, demographics and disposition
2	301, 302	Pivotal Pool: Includes all periods. Re-enrollers counted as distinct subjects for each enrollment.	Support pooled safety and efficacy analysis of pivotal studies
3	201, 301, 302	Comparison Pool: Includes all periods.	Support pooled safety analysis between study drug and comparators

Figure 5: Sample Pool Definitions

In addition to the definitions above, the SAP defines baseline values as the last value collected before the first exposure in each pool and treatment-emergent adverse event (TEAE) periods are based on the treatments received in each pool.

For subjects such as 02-003 and 02-004, their first exposure date for Pools 1 and 3 will be based on their participation in study MD-201, while their first exposure date for Pool 2 will be based their participation studies MD-301/302. Their baseline values are similarly impacted.

To support the pool-specific definitions for analysis, it is decided to use the new Integration structure classes.

IADSL

The ADSL metadata below in Table 5 includes the new integration variable POOL and two other new variables POOLN (the numeric version of POOL) and STUDIES (multiple STUDYID values in one string). The treatment and analysis period variables are aligned to the SAP definition for each pool.

Table 5: Sample ADSL Metadata

Variable	Label	Source/Derivation/Comments
USUBJID	Unique Subject Identifier	ISS.DM.USUBJID
POOL	Pool	For each SAP defined pool a subject belongs to, create a record. 'Overall' for overall pool if patient was enrolled in any study (101, 201, 301, 302, 320), 'Pivotal' for Pivotal pool if patient was enrolled in study 301 or 302. 'Comparison' for comparison pool if patient was enrolled in study 201, 302 or 320.
POOLN	Pool (N)	Map from POOL 'Overall'=1 'Pivotal'=2 'Comparison'=3
STUDIES	Study or Studies Identifier	A list, delimited by comma ',' , of ISS.DM.STUDYID values in the order of participation for each subject as applicable for the pool
TRT01P	Planned Treatment for Period 01	First treatment period in ISS.DM.ARM where ISS.DM.STUDYID is in the list of STUDIES
TR01SDT	Date of First Exposure in Period 01	The date of the first dose in the period described by TRT01P. Take EX records for the described period by checking ISS.EX.STUDYID and ISS.EX.EXTRT, take the minimum ISS.EX.EXSTDTC value converted to SAS® date9 format.
TR01EDT	Date of Last Exposure in Period 01	The date of the last dose in the period described by TRT01P. Take EX records for the described period by checking ISS.EX.STUDYID and ISS.EX.EXTRT, take the maximum ISS.EX.EXENDTC value converted to SAS date9 format.
AP01SDT	Period 01 Start Date	TR01SDT
AP01EDT	Period 01 End Date	The minimum of TR01EDT+7 and the next period start date-1
TRT02P	...	See TRT01P, apply for second exposure period
TR02SDT	...	See TR01SDT, apply for second exposure period
TR02EDT	...	See TR01EDT, apply for second exposure period
AP02SDT	...	TR02SDT
AP02EDT	...	The minimum of TR02EDT+7 and the next period start date-1
TRT03P	...	See TRT01P, apply for third exposure period

Variable	Label	Source/Derivation/Comments
TR03SDT	...	See TR01SDT, apply for third exposure period
TR03EDT	...	See TR01EDT, apply for third exposure period
AP03SDT	...	TR03SDT
AP03EDT	...	The minimum of TR03EDT+7 and the next period start date-1
TRT04P	...	See TRT01P, apply for fourth exposure period
TR04SDT	...	See TR01SDT, apply for fourth exposure period
TR04EDT	...	See TR01EDT, apply for fourth exposure period
AP04SDT	...	TR04SDT
AP04EDT	...	TR04EDT+7

Table 6 below shows sample ADSL records for the variables described in Table 5. There is a record for each subject and pool the subject belongs in. The variable STUDIES lists the studies included in each record, and all values on the record reflect pool-specific derivations. Additional variables such as demographics, population flags, covariates, and baseline variables are expected in a real study.

Table 6: Sample ADSL Data

ROW	USUBJID	POOLN	POOL	STUDIES	TRT01P	TR01SDT	TR01EDT	AP01SDT
1	MD-101-01-007	1	Overall	MD-101	MD 10mg	2000-02-01	2000-02-07	2000-02-01
2	MD-201-02-003	1	Overall	MD-201, MD-301, MD-320	MD 10mg	2000-08-10	2000-09-02	2000-08-10
3	MD-201-02-003	2	Pivotal	MD-301	MD 10mg	2001-08-21	2002-04-11	2001-08-21
4	MD-201-02-003	3	Comparison	MD-201, MD-301	MD 10mg	2000-08-10	2000-09-02	2000-08-10
5	MD-201-02-004	1	Overall	MD-201, MD-302, MD-320	SOC 20mg	2000-08-29	2000-09-24	2000-08-29
6	MD-201-02-004	2	Pivotal	MD-302	SOC 20mg	2001-09-06	2002-04-27	2001-09-06
7	MD-201-02-004	3	Comparison	MD-201, MD-302	SOC 20mg	2000-08-29	2000-09-24	2000-08-29

ROW	AP01EDT	TRT02P	TR02SDT	TR02EDT	AP02SDT	AP02EDT	TRT03P	TR03SDT
1(C)	2000-02-07	MD 20mg	2000-02-08	2000-02-10	2000-02-08	2000-03-17		
2(C)	2000-09-09	SOC 20mg	2000-09-10	2000-10-03	2000-09-10	2000-10-10	MD 10mg	2001-08-21
3(C)	2002-04-18							
4(C)	2000-09-09	SOC 20mg	2000-09-10	2000-10-03	2000-09-10	2000-11-02	MD 10mg	2001-08-21
5(C)	2000-10-01	MD 10mg	2000-10-02	2000-10-27	2000-10-02	2000-11-03	SOC 20mg	2001-09-06
6(C)	2002-05-01							
7(C)	2000-10-01	MD 10mg	2000-10-02	2000-10-27	2000-10-02	2000-11-03	SOC 20mg	2001-09-06

ROW	TR03EDT	AP03SDT	AP03EDT	TRT04P	TR04SDT	TR04EDT	AP04SDT	AP04EDT
1(C)								
2(C)	2002-04-11	2001-08-21	2002-04-18	MD 10mg	2002-05-13	2004-10-16	2002-05-13	2004-10-23
3(C)								
4(C)	2002-04-11	2001-08-21	2002-05-11					
5(C)	2002-04-27	2001-09-06	2002-05-01	MD 10mg	2002-05-02	2005-02-01	2002-05-02	2005-02-08
6(C)								
7(C)	2002-04-27	2001-09-06	2002-05-01					

IOCCDS

The ADAE metadata below in Table 7 includes the new integration variable POOLN merged in from ADSL. Other variables from ADSL are implicitly included and not relisted. A set of ADAE records is created for pools 2 and 3 only, intended to show that the producer may create records as needed. The variables STUDYID and AESEQ provide data point traceability back to the SDTM source records. The derivation of TRTP and TRTEMFL do not need to refer to pool definitions because the variables APxxSDT/APxxEDT merged in from ADSL are pool-specific, simplifying the derivation of this variable.

Table 7: Sample ADAE Metadata

Variable	Label	Source/Derivation/Comments
USUBJID	Unique Subject Identifier	ISS.AE.USUBJID
POOLN	Pool (N)	ADSL.POOLN only include records where POOLN=2,3
STUDYID	Study or Studies Identifier	ISS.AE.STUDYID
AESEQ	Sequence Number	ISS.AE.AESEQ
AEDECOD	Dictionary-Derived Term	ISS.AE.AEDECOD
AEBODSYS	Body System or Organ Class	ISS.AE.AEBODSYS
ASTDT	Analysis Start Date	Numeric version of ISS.AE.AESTDTC Note: AE start date derivation usually include imputation for partial dates, please see the OCCDS Implementation Guide for a more detailed example
TRTP	Planned Treatment	If APxxSDT<=ASTDT<=APxxEDT for xx=01 thru 04, then set TRTP=the corresponding TRTxxP variable. For example, if AP02SDT<=ASTDT<=AP02EDT, then set TRTP=TRT02P
TRTEMFL	Treatment Emergent Analysis Flag	Treatment emergent adverse events are defined as those that start within a treatment exposure period + 7 days or the day before the start of the next exposure period, whichever is less. If APxxSDT<=ASTDT<=APxxEDT for xx=01 thru 04, then set TRTEMFL='Y'.

Table 8 below shows sample ADAE records for the variables described in Table 7, omitting AEBODSYS. Records supporting pool 2 only include records from studies MD-301/302 while pool 3 records also include records from MD-301. Data point traceability is clear from SDTM variables and derivation traceability for TRTP and TRTEMFL would be clear from the inclusion of ADSL variables.

Table 8: Sample ADAE Data

USUBJID	POOLN	STUDYID	AESEQ	AEDECOD	ASTDT	TRTP	TRTEMFL
MD-201-02-003	2	MD-301	1	Epistaxis	2001-09-12	MD 10mg	Y
MD-201-02-003	2	MD-301	2	Hypotension	2002-04-19		
MD-201-02-003	3	MD-201	1	Headache	2000-08-10	MD 10mg	Y
MD-201-02-003	3	MD-201	2	Back pain	2000-09-11	SOC 20mg	Y
MD-201-02-003	3	MD-301	1	Epistaxis	2001-09-12	MD 10mg	Y
MD-201-02-003	3	MD-301	2	Hypotension	2002-04-19	MD 10mg	Y
MD-201-02-004	2	MD-302	1	Hypotension	2001-11-06	SOC 20mg	Y
MD-201-02-004	2	MD-302	2	Diarrhoea	2002-04-05	SOC 20mg	Y
MD-201-02-004	3	MD-201	1	Back pain	2000-09-20	SOC 20mg	Y
MD-201-02-004	3	MD-201	2	Epistaxis	2000-11-05		
MD-201-02-004	3	MD-302	1	Hypotension	2001-11-06	SOC 20mg	Y
MD-201-02-004	3	MD-302	2	Diarrhoea	2002-04-05	SOC 20mg	Y

Please note that the same AE record for subject 02-003 from MD-301 with AESEQ=2 is considered a TEAE for pool 3 but not for pool 2. SAP pool definitions can result in different values for TRTEMFL, ADY, TRTP, TRTA, APERIOD, and other variables for the same AE record. Having a set of records for each pool means standard OCCDS variables may be used, whereas pool-specific custom variables would need to be created if the same record was to support all pools.

IBDS

The ADLB metadata below in Table 9 includes the new integration variable POOL merged in from ADSL. A set of records is created for pools 2 and 3 only as defined in the SAP. The variables STUDYID and LBSEQ provide data point traceability back to SDTM. The value of AVISIT, ADY, ABLFL, and TRTP may differ depending on the pool, however as pool-specific ADSL variable values are merged in, the derivations remain simple.

Table 9: Sample ADLB Metadata

Variable	Label	Where Condition	Source/Derivation/Comments
USUBJID	Unique Subject Identifier		ADSL.USUBJID
POOLN	Pool (N)		ADSL.POOLN only include records where POOLN= 2, 3
STUDYID	Study Identifier		ISS.LB.STUDYID
LBSEQ	Sequence Number		ISS.LB.LBSEQ

Variable	Label	Where Condition	Source/Derivation/Comments
PARAM	Parameter		ISS.LB.LBTEST (ISS.LB.LBSTRESU) Note: In practice PARAM would also include LBSPEC and/or LBMETHOD, etc. (e.g. Urine Glucose)
AVAL	Analysis Value		ISS.LB.LBSTRESN
ADT	Analysis Date		Numeric version of ISS.LB.LBDTC
ADY	Analysis Day		ADT - TR01SDT + (ADT>=TR01SDT)
AVISIT	Analysis Visit	POOLN EQ 2	Proper case of ISS.LB.VISIT
-	-	POOLN EQ 3	If ABLFL=Y then 'Baseline', If TRTP is missing then 'Not in Pool', Otherwise if TRTP is not missing then map using ADY 2-30='Days 2-30' 31-150='Days 31-150' 151-380='Days 151-380' 381-500='Days 381-500'
ABLFL	Baseline Record Flag		Set to "Y" for last pre-dose record where ADY<=1 and ISS.LB.LBTPT=PREDOSE (Note: Time point is not shown in this example)
TRTP	Planned Treatment		If APxxSDT<=ASTDT<=APxxEDT for xx=01 thru 04, then set TRTP=the corresponding TRTxxP variable. For example, if AP02SDT<=ASTDT<=AP02EDT, then set TRTP=TRT02P

Table 10 below shows sample ADLB records for the variables described in Table 9. Records supporting pool 2 only include records from studies MD-301/302 while pool 3 records also include records from MD-301. Data point traceability is clear from SDTM variables and derivation traceability for timing related variables would be clear from the inclusion of ADSL variables.

Table 10: Sample ADLB Data

USUBJID	POOLN	STUDYID	LBSEQ	PARAM	AVAL	ADT	ADY	AVISIT	ABLFL	TRTP
MD-201-02-003	2	MD-301	1	Glucose	96	2001-08-21	1	Baseline	Y	MD 10mg
MD-201-02-003	2	MD-301	2	Glucose	87	2001-08-29	9	Week 1		MD 10mg
MD-201-02-003	3	MD-201	1	Glucose	98	2000-08-10	1	Baseline	Y	MD 10mg
MD-201-02-003	3	MD-201	2	Glucose	78	2000-08-17	8	Days 2-30		MD 10mg
MD-201-02-003	3	MD-301	1	Glucose	96	2001-08-21	377	Days 151-380		MD 10mg
MD-201-02-003	3	MD-301	2	Glucose	87	2001-08-29	385	Days 381-500		MD 10mg
MD-201-02-004	2	MD-302	1	Glucose	71	2001-09-06	1	Baseline	Y	SOC 20mg
MD-201-02-004	2	MD-302	2	Glucose	75	2001-09-13	8	Week 1		SOC 20mg

USUBJID	POOLN	STUDYID	LBSEQ	PARAM	AVAL	ADT	ADY	AVISIT	ABLFL	TRTP
MD-201-02-004	3	MD-201	1	Glucose	79	2000-08-29	1	Baseline	Y	SOC 20mg
MD-201-02-004	3	MD-201	2	Glucose	85	2000-09-06	9	Days 2-30		SOC 20mg
MD-201-02-004	3	MD-302	1	Glucose	71	2001-09-06	374	Days 151-380		SOC 20mg
MD-201-02-004	3	MD-302	2	Glucose	75	2001-09-13	381	Days 381-500		SOC 20mg

Please note that conceivably, the variable BASETYPE could be used for the same purpose as POOLN in this dataset for ABLFL and AVISIT derivations, however BASETYPE would not serve as a key for merging in pool-specific variables from ADSL. The variable BASETYPE remains available if there are multiple baselines defined within a pool.

CONCLUSION

There are more IADSL standard variables defined in the ADaM Integration standards document than the ones presented in this paper. On the other hand, there are no new standard variables for the IBDS and IOCCDS structures other than those merged in from the ADSL dataset (IADSL class). When merged in, the new IADSL class variables have an impact on the usage of variables such as ABLFL, BASE, OCCzzFL, and others. In order to promote understanding of these changes, and permit compliance rules around them, new structure classes are necessary.

Existing ADaM classes may be used for integration, and as seen through planning in advance, when similar study and dataset designs are utilized with the same controlled terminology, integration efforts are greatly reduced.

For the challenge of more complex integrations, namely when analyses vary by pool, the new IADSL/IBDS/IOCCDS structures build this variability vertically into the analysis datasets. The advantages of this approach are the reusability of the existing ADSL class variables pertaining to subjects within pool rather than just subjects, and the ability to subset all the ADaM datasets by pool and quickly understand the experience of a subject within a pool.

REFERENCES

[1] FDA. “Guidance for Industry Integrated Summaries of Effectiveness and Safety: Location Within the Common Technical Document.” Published April 2009. Available at https://www.fda.gov/downloads/drugs/guidances/ucm136174.pdf.

RECOMMENDED READING

- *Analysis Data Model Implementation Guide version 1.1*
- *Analysis Data Model Structure for Occurrence Data (OCCDS) version 1.0*
- *CDISC Define-XML Specification Version 2.0*

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Wayne Zhong
Accretion Softworks
wayne@asoftworks.com

Kimberly Minkalis
The Griesser Group
kminkalis@gmail.com

Deborah Bauer
Sanofi
Deborah.Bauer@sanofi.com